Discovering New Strong Gravitational Lenses in the DESI Legacy Imaging Surveys

X. Huang[1] , C. Storfer[1] , A. Gu[2,3], V. Ravi[4], A. Pilon[1], W. Sheu[2,3], R. Venguswamy[5], S. Banka[3], A. Dey[6] , M. Landriau[7] ,
D. Lang[8,9,10] , A. Meisner[6] , J. Moustakas[11], A. D. Myers[12], R. Sajith[2,3], E. F. Schlafly[13] , and D. J. Schlegel[7]

[1] Department of Physics & Astronomy, University of San Francisco, San Francisco, CA 94117-1080, USA; xhuang22@usfca.edu
[2] Department of Physics, University of California, Berkeley, Berkeley, CA 94720, USA
[3] Department of Electrical Engineering & Computer Sciences, University of California, Berkeley, Berkeley, CA 94720, USA
[4] Department of Computer Science, University of San Francisco, San Francisco, CA 94117-1080, USA
[5] Department of Computing, Data Science, and Society, University of California, Berkeley, Berkeley, CA 94720, USA
[6] NSF's National Optical–Infrared Astronomy Research Laboratory, 950 N. Cherry Avenue, Tucson, AZ 85719, USA
[7] Physics Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
[8] Dunlap Institute, University of Toronto, Toronto, ON M5S 3H4, Canada
[9] Department of Astronomy & Astrophysics, University of Toronto, Toronto, ON M5S 3H4, Canada
[10] Perimeter Institute for Theoretical Physics, Waterloo, ON N2L 2Y5, Canada
[11] Department of Physics and Astronomy, Siena College, 515 Loudon Road, Loudonville, NY 12211, USA
[12] Department of Physics & Astronomy, University of Wyoming, 1000 E. University, Dept. 3905, Laramie, WY 82071, USA
[13] Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550-9234, USA

Abstract

We have conducted a search for new strong gravitational lensing systems in the Dark Energy Spectroscopic Instrument Legacy Imaging Surveys' Data Release 8. We use deep residual neural networks, building on previous work presented by Huang et al. These surveys together cover approximately one-third of the sky visible from the Northern Hemisphere, reaching a z-band AB magnitude of $\sim$22.5. We compile a training sample that consists of known lensing systems as well as non-lenses in the Legacy Surveys and the Dark Energy Survey. After applying our trained neural networks to the survey data, we visually inspect and rank images with probabilities above a threshold. Here we present 1210 new strong lens candidates.

Unified Astronomy Thesaurus concepts: Strong gravitational lensing (1643); High-redshift galaxies (734); AGN host galaxies (2017); Galaxies (573); Galaxy clusters (584); Galaxy groups (597); Quasars (1319)

1. Introduction

Strong gravitational lensing systems are a powerful tool for astrophysics and cosmology. They have been used to study how dark matter is distributed in galaxies and galaxy clusters (e.g., Kochanek 1991; Koopmans & Treu 2002; Bolton et al. 2006; Koopmans et al. 2006; Vegetti & Koopmans 2009; Tessore et al. 2016; Monna et al. 2017; Jauzac et al. 2018; Shajib et al. 2019; Meneghetti et al. 2020), and are uniquely suited to probe dark matter substructure beyond the local universe (e.g., Vegetti et al. 2014, 2018; Ritondale et al. 2019; Diaz Rivero & Dvorkin 2020). Furthermore, by modeling galaxy-scale strong lenses as the sum of luminous and dark components, the mass-to-light ratio and inner density profile of the dark matter halo can be simultaneously constrained (e.g., Auger et al. 2010; Sonnenfeld et al. 2019; Shajib et al. 2020). Such measurements for a large number of lensing systems over a wide range of redshifts make it possible to study the structural evolution of massive elliptical galaxies, and possibly in the future, lenses of any Hubble type (e.g., Sonnenfeld et al. 2015; Nightingale et al. 2019). For nearby strong lensing galaxies, extragalactic tests of general relativity can be performed by combining lens modeling with spatially resolved stellar kinematic observations (Collett et al. 2018).

For redshifts beyond the range that lensing galaxies can typically probe, together with high-resolution imaging (from the Hubble Space Telescope, adaptive optics, or the James Webb Space Telescope in the near future), strong lensing as a cosmic telescope magnifies spectral (e.g., Cornachione et al. 2018) and spatial features (e.g., Marshall et al. 2007; Patrício et al. 2019; Vanzella et al. 2020) of the lensed distant galaxies,

providing the only way to study the morphology and internal structures of galaxies at sub-kiloparsec scales at high redshifts that can extend to $z > 2$.

Recent measurements of the Hubble constant H_0 span a range of $\sim$10% (e.g., Abbott et al. 2017, 2018; Freedman et al. 2019, 2020; Riess et al. 2019; Wong et al. 2020; Choi et al. 2020; Khetan et al. 2020; Philcox et al. 2020; Planck Collaboration et al. 2020), and significant tension remains between predictions for H_0 based on early-universe observables and direct late-universe measurements (e.g., Verde et al. 2019). Multiply lensed supernovae (SNe) are ideal for measuring time delays and H_0 because of their well-characterized light curves, and in the case of Type Ia, with the added benefit of standardizable luminosity (Refsdal 1964; Oguri & Marshall 2010; Treu 2010). In recent years, strongly lensed supernovae, both core-collapse (Kelly et al. 2015; Rodney et al. 2016) and Type Ia (Quimby et al. 2014; Goobar et al. 2017), have been discovered. Time-delay H_0 measurements from multiply imaged supernovae can therefore be an important independent approach to address the discrepancy between H_0 measured locally and the value inferred from the cosmic microwave background (e.g., Goldstein & Nugent 2017; Goldstein et al. 2018, 2019; Pierel & Rodney 2019; Wojtak et al. 2019; Suyu et al. 2020).

Furthermore, time delay H_0 measurements are a powerful complement to other independent measurements of the equation of state of dark energy (e.g., Linder 2011; Treu & Marshall 2016). Beyond the flat ΛCDM cosmological model, by combining strong lensing time delay of multiply imaged time-varying sources and SNe Ia distance measurements, one can determine H_0 in a model-independent way and measure the
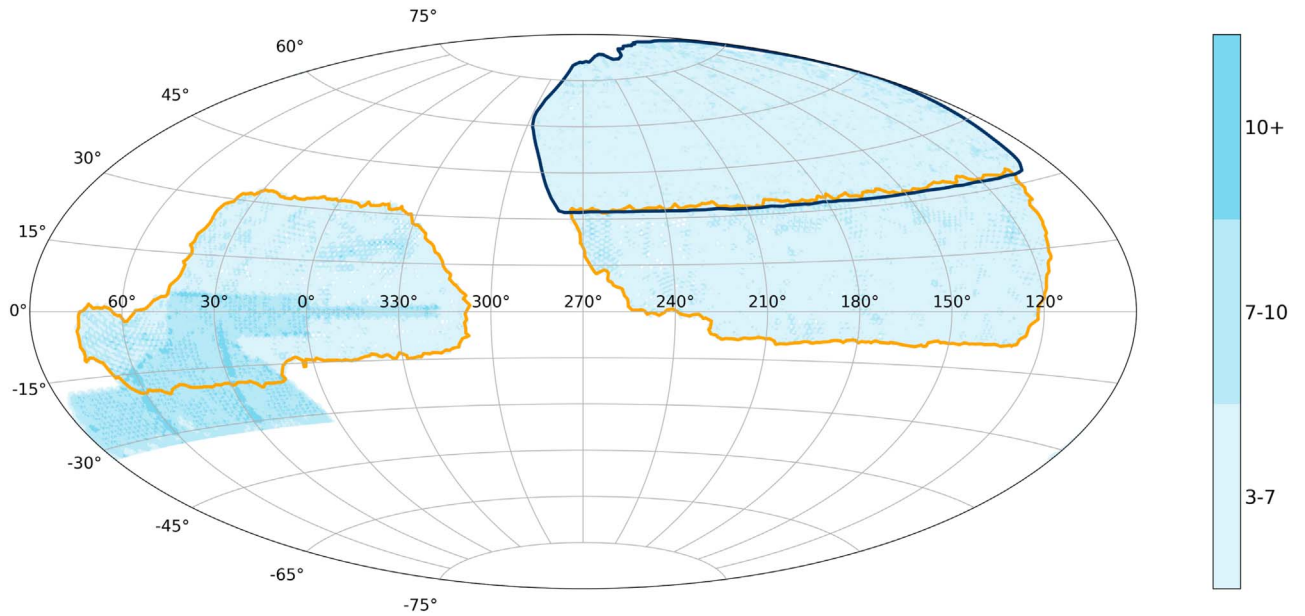
Figure 1. The DESI Legacy Imaging Surveys footprint in an equal-area Aitoff projection in equatorial coordinates. The blue and gold borders approximately outline the north (coinciding with MzLS/BASS) and south (residing within DECaLS) regions of the spectroscopic survey, respectively. Slightly above $\delta = 32°$, there is a small amount of overlap between the imaging surveys. Patches with different shades of blue indicate the depth in the z band: light for between three and seven passes, medium for between seven and ten, and dark for more than ten. Note that DECaLS includes the DES footprint, but has incomplete coverage below $\delta \approx -32°$ in Data Release 8.

spatial curvature (e.g., Li et al. 2018; Collett et al. 2019; Taubenberger et al. 2019), and test the Friedmann–Lemaître–Robertson–Walker metric (Räsänen et al. 2015).

For many of these analyses, the available sample sizes of confirmed strong lenses are a major limiting factor. In the last few years, several groups have used convolutional neural networks (CNNs) to search for strong lensing systems in photometric surveys including, in increasing sky coverage, CFHTLS (Jacobs et al. 2017), KiDS (Petrillo et al. 2017, 2019; Li et al. 2020), DES (Jacobs et al. 2019a, 2019b), and Pan-STARRS (Canameras et al. 2020).

Data release 8 (DR8) of the DESI Legacy Surveys[14] (Dey et al. 2019), for which at least the z band is observed with a 4 m telescope, covers ~14,000 deg², three times the size of the DES footprint. In Huang et al. (2020, hereafter H20), we identified hundreds of new strong lenses in the Legacy Surveys Data Release 7 (DR7) by using a residual neural network (ResNet). In this paper, building on H20, we have significantly improved the efficiency of the neural network and report the discovery of new strong lensing systems over a wide range of redshifts in DESI Legacy Surveys DR8.

This paper is organized as follows. A brief description of the Legacy Surveys is given in Section 2. In Section 3, we describe our methodology, including the improvements we have made on H20. In Section 4, we show the inference results and present our best candidates for strong lensing systems. We discuss our results in Section 5, and conclude in Section 6.

2. Observations

The Legacy Imaging Surveys consist of three projects: the Dark Energy Camera Legacy Survey (DECaLS), observed by the Dark Energy Camera (DECam; Flaugher et al. 2015) on the

4 m Blanco telescope at the Cerro Tololo Inter-American Observatory; the Beijing–Arizona Sky Survey (BASS), by the 90Prime camera (Williams et al. 2004) on the Bok 2.3 m telescope owned and operated by the University of Arizona and located on Kitt Peak; and the Mayall z-band Legacy Survey (MzLS), by the Mosaic3 camera (Dey et al. 2016) on the 4 m Mayall telescope at Kitt Peak National Observatory. Together they cover ~14,000 deg² of the extragalactic sky visible from the Northern Hemisphere with at least three passes in each of the three bands, grz. The 5σ z-band median limiting AB magnitude is 22.5 mag for galaxies with an exponential disk profile with $r_{\mathrm{half}} = 0\rlap{.}''45$.

The combined survey footprint is split into two contiguous areas by the Galactic plane. DECaLS covers the ~9000 deg² subregion at $\delta \lesssim +32°$ of the Legacy Surveys, while BASS/MzLS covers the ~5000 deg² northern subregion. Figure 1 shows the different regions in the the Legacy Surveys footprint and the depth of the z-band observation.

For DECaLS (gold outline in Figure 1), the delivered image quality has FWHM of approximately $1\rlap{.}''29$, $1\rlap{.}''18$, and $1\rlap{.}''11$ for g, r, and z bands, respectively. For the $\delta \gtrsim +32°$ footprint (blue outline in Figure 1) of the Legacy Surveys, MzLS has obtained images in the z-band that complemented the BASS g- and r-band observations in the same subregion, with median delivered image quality of approximately $1\rlap{.}''61$, $1\rlap{.}''47$, and $1\rlap{.}''01$ for g, r, and z bands, respectively.

The Legacy Surveys used *The Tractor* package (Lang et al. 2016) as a forward-modeling approach to perform source extraction on pixel-level data. *The Tractor* takes as input the individual images from multiple exposures in multiple bands, with different seeing in each. After source detection, the point source ("PSF") and spatially extended ("REX," round exponential galaxy) models are computed for every source, and the better of these two is used when deciding whether to keep the source. The spatially extended sources (REX) are

---

[14] http://www.legacysurvey.org/
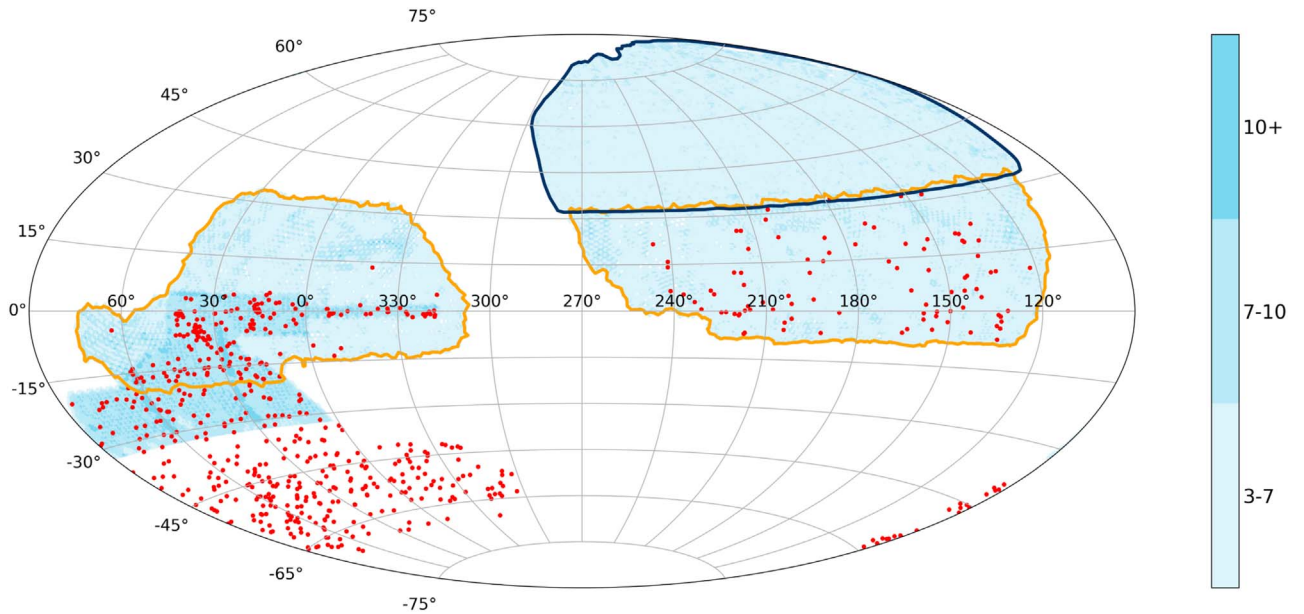
Figure 2. Previously known lenses or lens candidates in our training sample shown as red dots, against the background of the depth map of Legacy Surveys DR8 (see the caption for Figure 1). The lenses south of the DESI spectroscopic footprint (gold outline) are from DES.

further classified according to whether χ^2 is improved by 9 by treating it as a de Vaucouleurs (DEV) profile, an exponential (EXP) profile, or a composite of DEV + EXP, or COMP.[15] The same light profile (EXP, DEV, or COMP) is consistently fit to all images in order to determine the best-fit source shape parameters and photometry.

The categories of DEV and COMP indicate the classification of elliptical galaxies. Given that the vast majority of lensing events are caused by massive early-type galaxies, we decided to first target objects with DEV and COMP classifications, and then REX, which tend to be smaller and/or fainter galaxies.

3. The Training Sample and ResNets

Deep CNNs and their variations have been shown to be highly effective in image recognition. In recent years, this technique has been successfully applied to recognize instances of strong lenses in simulations (e.g., Metcalf et al. 2019, and references therein). As mentioned in Section 1, several groups have searched for and found strong lenses in existing imaging surveys. In all these efforts, at least the positive examples (lenses) in the training samples were constructed from simulated lenses, typically of the order of $\mathcal{O}(10^5)$. This is because the number of known lenses, of the order of several hundred, is thought to be too small to effectively train CNN models. Building on H20, we continue to use only *observed* data for lenses and non-lenses in our training sample. In this section (and Section 4), we show we can train deep neural networks with a much smaller sample and far fewer positive examples and achieve comparable if not superior results. In Section 3.1, we describe our training sample. We show the training results using the ResNet from Lanusse et al. (2018) in Section 3.2. Finally in Section 3.3, we present an improved neural network model.

3.1. Training Sample

The Master Lens Database[16] (Moustakas et al. 2012), which contains hundreds of lensing events discovered prior to 2016, provided the initial list for the lens training sample. We have since added several hundred more lenses and lens candidates from more recent publications (Carrasco et al. 2017; Diehl et al. 2017; Jacobs et al. 2017, 2019a, 2019b; Pourrahmani et al. 2018; Sonnenfeld et al. 2018; Wong et al. 2018; H20). Initially our primary goal was to find lenses in DECaLS, part of which was observed by DES. Therefore in total we have identified 632 previously known lenses or lens candidates in DECaLS and DES, to be used in our training sample. For the lenses in the DES footprint, we only include *grz* bands. We also assemble $\sim$21,000 non-lens image cutouts from DECaLS and DES, all with at least three passes in each of the *grz* bands, a *z*-band mag <20.0, and typed as DEV or COMP, randomly distributed in the footprint. Given that on average we expect one strong lens in $\mathcal{O}(10^4)$ galaxies (e.g., Collett 2015; Jacobs et al. 2019b) incidental inclusion of a lens or two in these randomly selected galaxies is not a significant concern.

In the training sample of H20, we found that the images for the lenses tend to be much deeper than those for the non-lenses. This led to the neural net during the inference stage preferentially assigning high probabilities to images with deeper observations whether they are lenses or not. To correct for this bias, given that many (359) of our lenses in the training sample are from DES south of $\delta = -18°$ with deeper observations (see Figure 2), we have included in the non-lens sample 5000 random cutout images from the same region.

As with H20, included in the non-lens sample are cutouts selected by eye so as to cover as many non-lens configurations as possible, especially cases that can potentially be confused by the neural net. These include spiral galaxies of different sizes and spiral arm configurations, elliptical galaxies, galaxy groups, images having objects with different colors (typically a blue galaxy next to a red one), cosmic rays appearing in

---
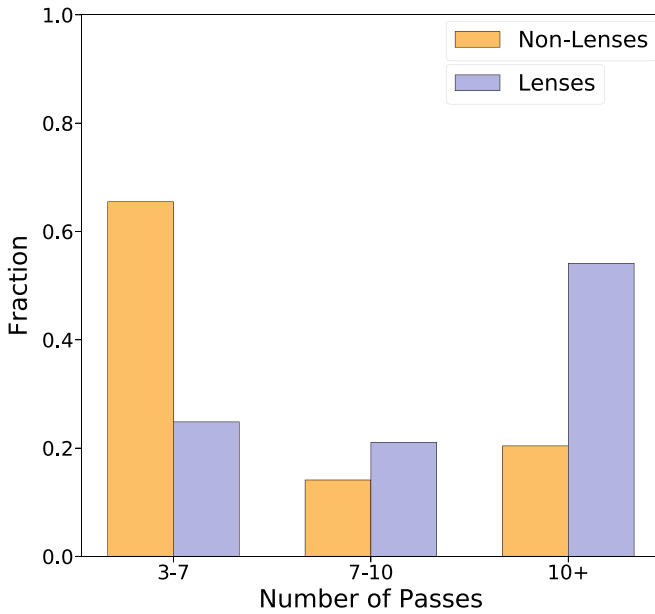
Figure 3. The violet and yellow columns show the fractions of lenses and non-lenses in the training sample, respectively, for the three bins of z-band depth.

different bands (some of which have curved trajectories), unusual arrangements of galaxies or stars, and finally certain data reduction artefacts.

The distribution of the lenses and non-lenses in our training sample is shown in Figure 3. While fractionally there are still more non-lenses in the shallowest bin and more lenses in the deepest bin, overall the disparity between the relative proportions of lenses and non-lenses in each depth bin is much improved compared with the training sample in H20.

3.2. Residual Neural Networks

We use the ResNet model of Lanusse et al. (2018, hereafter L18), after re-implementing it in TensorFlow.[17] We have left their architecture and hyperparameters unchanged (for details, see Section 3.3 of L18), except that we double the batch size to 256. The lens and non-lens images in the training sample are cutouts with a dimension of 101×101 pixels, following the specification in the Lens Challenge (Metcalf et al. 2019).

We split the training sample into training and validation *sets*, with a ratio of 7:3. We then train the ResNet on Google Colab[18] using a graphics processing unit (NVIDIA Tesla v100). The 120 epochs of training took 4 hr.

The ResNet attempts to minimize the cross-entropy loss function:

$$-\sum_{i=1}^{N} y_i \log \hat{y}_i + (1 - y_i)\log(1 - \hat{y}_i) \qquad (1)$$

where y_i is a label for the ith image (1 for lens and 0 for non-lens), and $\hat{y}_i \in [0, 1]$ is the model-predicted probability.

While the loss function is monitored during the training process to determine the point of termination, the overall performance of the trained model is typically assessed by the

receiver operating characteristic (ROC) curve. The ROC curve shows the true positive rate (TPR) versus the false positive rate (FPR) for the validation set, where P(ositive) indicates a lens and N(egative) a non-lens. With the definitions true positive (TP) = correctly identified as a lens, false positive (FP) = incorrectly identified as a lens, true negative (TN) = correctly rejected, and false negative (FN) = incorrectly rejected,

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{P}} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

and

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{N}} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}.$$

The curve is generated by gradually increasing the threshold probability for a positive identification from 0 to 1. Random classifications will result in a diagonal line in this space with an area under the ROC curve (or AUC) equal to 0.5. For a perfect classifier, AUC = 1.

In Figure 4, left panel, we show how the cross-entropy loss functions vary as training progresses. For the validation set, we show the value at every epoch. For the training set, the cross-entropy was reported for every step, which we have boxcar-smoothed with a window size of 57. This is because the training set has a total of 14,725 images, so with a batch size of 256 images it takes approximately 57 steps to complete one full training epoch. Figure 4 shows that the AUC for the validation set has plateaued well within the 120 epochs of training. We achieve an AUC of 0.992 for the validation set (Figure 4, right panel). This is a significant improvement over an already high AUC of 0.977 from H20.

3.3. Improvement on the L18 Model

We have experimented with a variety of ways to improve on the model in L18, including transfer learning and domain adaptation (e.g., Tzeng et al. 2017), among other techniques. We will provide the full results of the comparisons from these different approaches in a future publication.

So far we have run inference and visually inspected the results for one of the variants. In this modification on the original L18 model was the addition of "shielding" layers, inspired by the InceptionNet architecture of Szegedy et al. (2014). These "shields" are 1×1 convolutional layers inserted between every three blocks of the L18 architecture (see their Figure 4), so named because they have the effect of reducing dimensionality and mitigating the exponential increase in computational complexity present in the original architecture. With appropriate adjustments to the number of channels in the shielding layers, we reduce the number of trainable parameters by a factor of 50 (from 3 million to 60 thousand), thereby shortening the training time by 17%. Moreover, the validation AUC has increased from 0.992 (using the original L18 model; Section 3.2) to 0.997. Thus the reduction in model complexity does not appear to have an adverse impact on performance, and in fact has improved it. This is likely because the problem at hand (to tell lenses apart from non-lenses), although complex, does not require a large number of dimensions in the underlying latent space. The addition of "shielding" layers compresses dimensionality by more than an order of magnitude, forcing the network to learn only the most salient features. For example, in the final block of the architecture in L18 (see

---

[17] https://www.tensorflow.org/
[18] https://colab.research.google.com/
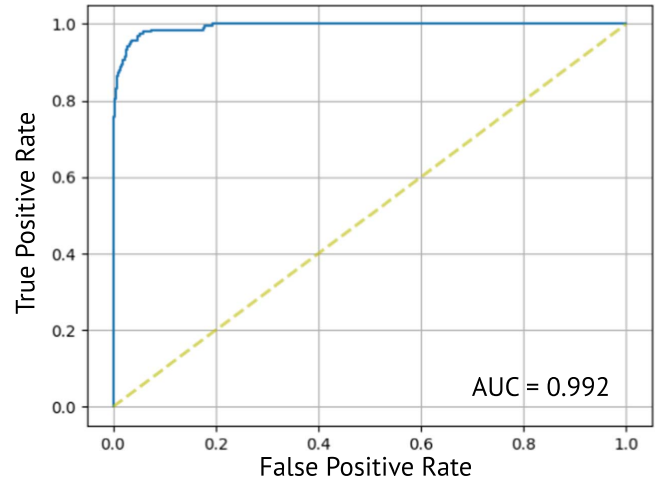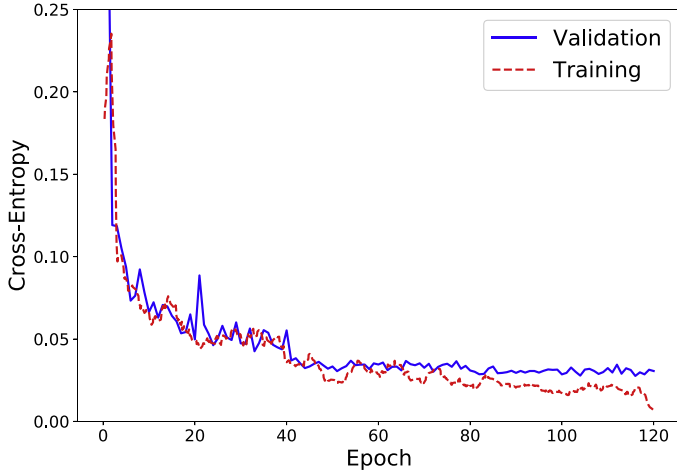
Figure 4. Left: the cross-entropy loss functions for the training and validation sets over 120 epochs. Right: the receiver operating characteristic curve for the validation set with the area under the curve = 0.992 for the last model after 120 epochs of training.

their Figure 4) we experimented with reducing the output from 512 channels to 256, 128, 64, 32, and 16 channels. We find that "shields" that keep the output to 32 channels perform the best.

In Section 4, we will show lens candidates from both the original model in L18 and the "shielded" model (the one with 32 output channels), to achieve greater completeness for the lens search in DR8 and to demonstrate that a different neural network model can identify new lens candidates.

4. Results

In this section we present the lens candidates. In Section 4.1, we present all the candidates found by using the ResNet model of L18, specifically: Section 4.1.1, candidates that are DEV or COMP in DECaLS; Section 4.1.2, candidates that are DEV or COMP found in BASS/MzLS; and Section 4.1.3, candidates that are typed as REX in DECaLS and MzLS. In Section 4.2, we show candidates that are found with the "shielded" model (see Section 3.3). To determine the probability threshold for human inspection, we consult the precision-recall curve (PRC), where precision = TP/(TP + FP) and recall = TP/(TP + FN), which is the same as TPR (Section 3.2). The PRC for the validation set, with probability threshold values marked, is shown in Figure 5.

We recognize that different terms have been used for the same quantities. To avoid confusion, in this paper:
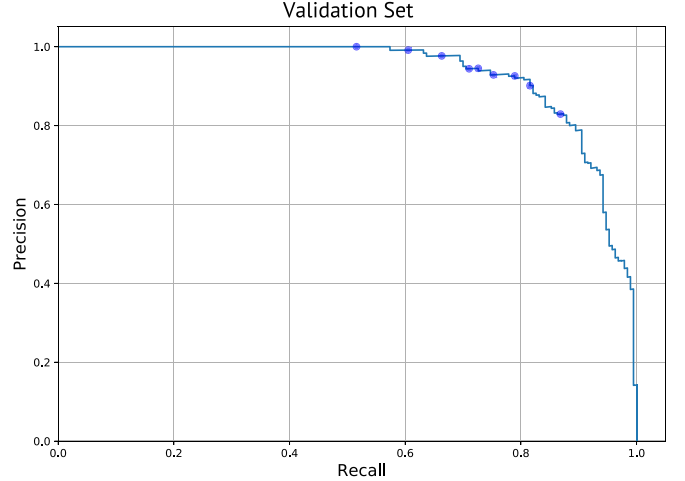
$$recall = TPR = completeness$$

and

$$precision = purity.$$

This redundancy in terminology in part stems from fairly standard usage (e.g., recall or TPR depending on the context) and in part from the difference in terminology between machine learning and astrophysics (recall or completeness, precision or purity).

While ideally we would like to identify all the lenses that are discoverable in the data set, there is a ceiling to the number of images that can be inspected in a reasonable amount of time. We choose the threshold of 0.1 because it seems to be a reasonable compromise between purity (precision) and completeness (recall). Keep in mind that the PRC provides completeness and purity for the validation set. For deployment

Figure 5. The precision-recall curve for the validation set. The blue points from left to right correspond to probability threshold values from 0.9 to 0.1 with an interval of 0.1.

on the whole data set, it is not possible to determine the completeness without inspecting the entire data set, which is infeasible. We will address the question of completeness in the context of comparing the results of different neural network models in a future publication (see Section 3.3). Since our training sample has a lens to non-lens ratio ($\sim$1 in 33) that is much higher than expected for the data set as a whole ($\sim$1 in 10^4), we estimate the expected purity for deployment at our chosen probability threshold of 0.1 in the following way. Given the 7:3 split between training and validation, there are approximately $N_l = 190$ lenses and $N_{nl} = 6300$ non-lenses in the validation set. The number of non-lenses misclassified as lenses is then $\sim$33 ($= N_l \times r \times \frac{1-p}{p}$), where r ($= 0.87$) and p ($= 0.83$) are the recall (or completeness) and precision (or purity), respectively. The fraction of non-lenses that are misclassified as lenses is $33/N_{nl} \approx 0.00052$. With the expectation of one strong lens in $\mathcal{O}(10^4)$ galaxies, this translates to a purity of 1 in 52, or 1.9%. We will refer to all cutout images with probabilities above this threshold as the ResNet "recommendations." Below, through human inspection, we will compare the percentage of lens candidates relative to the "recommendations" with this estimated purity for deployment.
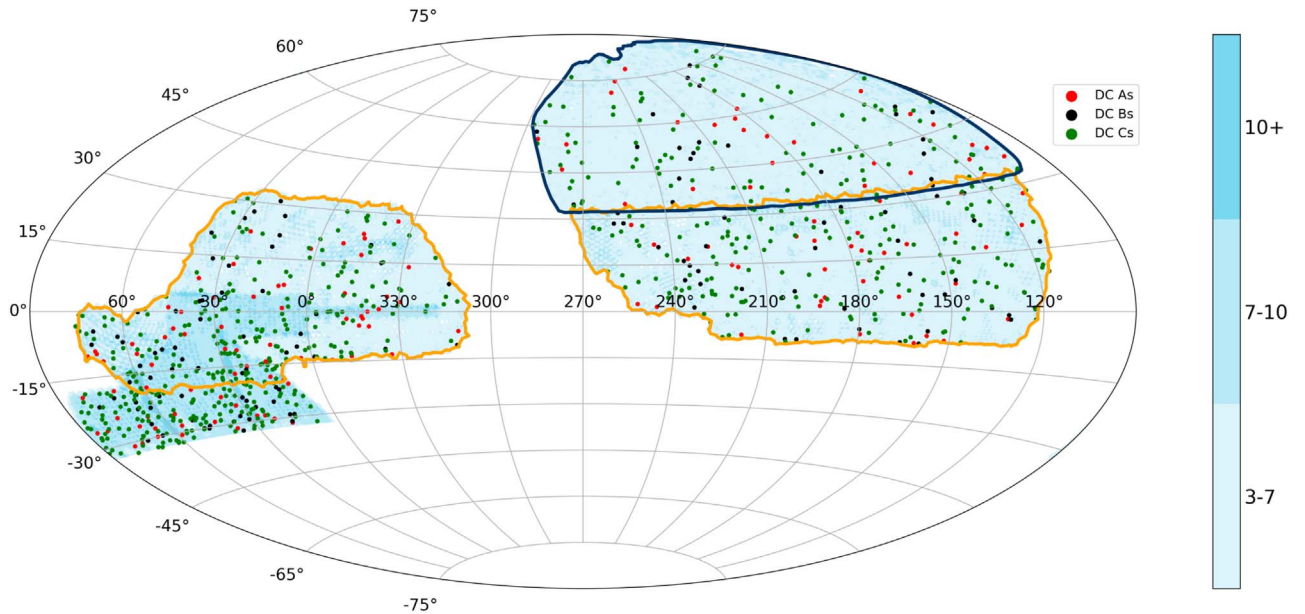
Figure 6. The new candidate lensing systems typed as DEV and COMP by *The Tractor* in the DECaLS and BASS/MzLS regions (see Figure 1 caption) are shown as red (Grade A), black (Grade B), and yellow (Grade C) circles.

(Note that in H20, we used the term "human inspection efficiency" for this quantity.)

Throughout this section, all objects we run inference on have $\geqslant 3$ passes for all three bands and z-band mag <20.0. For the ensuing human inspection, we follow this process. Coauthors S.B., A.G., A.P., V.R., C.S., W.S., and R.V. make the "first pass" selections, according to these criteria, erring on the generous side: small blue galaxy/galaxies (red galaxies are rare but certainly acceptable) next to the red galaxy/galaxies at the center that

1. are typically 1″–5″ away
2. have low surface brightness
3. curve toward the red galaxy/galaxies
4. have counter/multiple images with similar colors (especially in Einstein cross-like configuration)
5. are elongated (including semi- or nearly full circles).

Typically, most candidates do not have all these characteristics. In general, the greater the number of characteristics listed above that an image has, the higher they are ranked by humans. For the "second pass," coauthors X.H. and A.D. examine all "first pass" selections and each assign an integer score between 1 and 4. These two scores are averaged. We assign a letter grade according to the average, using the following criteria. For the third criterion below, and for the rest of the paper, we define the *angular scale* of a candidate system as the angular separation between the lens and the most prominent putative arc. Note that this can be somewhat different from (and in the case of a single large tangential arc, tends to be larger than) the extent of the critical curve (e.g., Narayan & Bartelmann 1996; Kneib & Natarajan 2011).

1. $\geqslant 3.5$: Grade A. We have a high level of confidence in these candidates. Many of them have one or more prominent arcs, usually blue. The rest have one or more clear arclets, sometimes arranged in multiple-image configurations with similar colors (again, typically blue). However, there are clear cases with red arcs.

2. $= 3.0$: Grade B. They have similar characteristics to the Grade A's. For the cutout images where there appear to be giant arcs they tend to be fainter than those for the Grade A's. Likewise, the putative arclets tend to be smaller and/or fainter, or isolated (without counter images).

3. $= 2.5$ or 2.0: Grade C. They generally have features that are even fainter and/or smaller than what is typical for Grade B candidates, but are nevertheless suggestive of lensed arclets. Counter images are often not present or indiscernible. In a number of cases, the angular scales of the candidate systems are comparable to or only slightly larger than the seeing. Therefore, for some of these candidates, higher spatial resolution or deeper data would be required to attain a higher level of certainty.

For Grade B and C candidates, we have included a small percentage of cases where it is difficult to judge whether it is a lensing event versus a coincidental placement of galaxies, a spiral galaxy, a ring galaxy, or tidal features associated with galaxy interactions.

4.1. Lens Candidates from the L18 ResNet

Below we present all the strong lens candidates found by using the ResNet model in L18.

4.1.1. Candidates from DEV and COMP in DECaLS

Searching for strong lenses among the DEV and COMP objects in the DECaLS region was originally our primary goal. Our training sample is selected from the same region (see Figure 2). We deploy our model on ~10 million cutouts centered on galaxies typed as DEV or COMP. With the probability threshold set at 0.1, in total we have examined 22,879 ResNet recommendations.

We have found 115 Grade A, 110 Grade B, and 501 Grade C candidates. The locations of these candidates in the sky are shown in Figure 6. In total, we have identified 726 candidates,
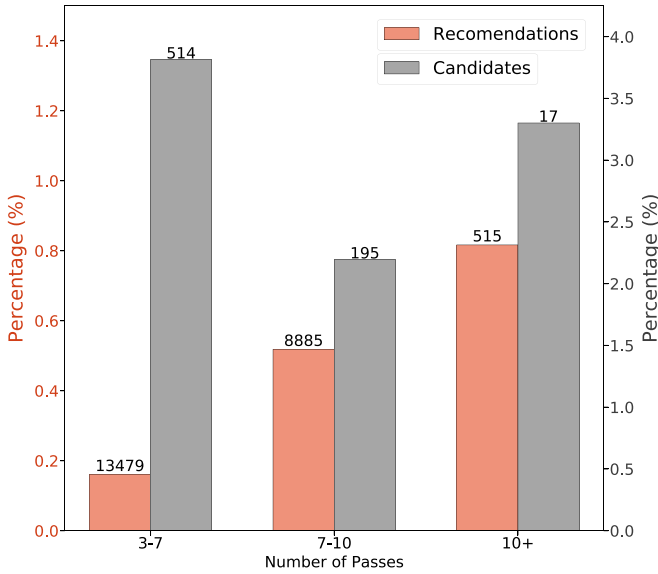
Figure 7. The orange columns (left *y*-axis) show the percentages of objects given a greater than 0.1 probability by our ResNet model (or "recommendations") for the three bins of *z*-band depth. The gray columns (right *y*-axis) show the percentages of ResNet recommendations that are selected as lens candidates through human inspection, or the purity. The number of recommendations or candidates for each bin is shown atop the corresponding column.

achieving a purity of approximately 1 in 31 ResNet recommendations.

We now briefly discuss the purity of the ResNet results thus far, since this is the primary data set in which we originally planned to search for lenses.

In H20, we noted that due to the composition of our training sample (comparatively smaller number of non-lens images with deep observations), the neural net showed a preference for images with deep observations, whether they contain lensing systems or not. For the inference results in this paper, Figure 7 shows how (a) the percentage of the ResNet recommendations relative to the objects and (b) the percentage of candidates (determined by human inspection) relative to recommendations depend on the observational depth (see Figure 1). For the three depth bins, the percentages of lens candidates relative to the neural net recommendations are similar, approximately between 2.2% and 3.8%. This indicates that the neural net now makes recommendations largely free of bias with regard to depth. This is consistent with our expectation based on the composition of the training sample used in this paper. The orange columns show that 0.82% of the objects in the bin for 10+ passes receive probability >0.1 ("recommendations"), five times the value of 0.16% in the bin for 3–7 passes. This trend in the ResNet recommendations indicates that, not surprisingly, there are more lenses to be discovered in deeper images. In fact, approximately one in 16,337, 8795, 3710 galaxies is a lens, from the shallowest to the deepest bin, assuming 100% completeness. These values are consistent with the expectation of one strong lens in $\mathcal{O}(10^4)$ galaxies.

Overall, our ResNet model achieves a purity of 3.2%, broadly consistent with our estimation of 1.9% (see the introduction to Section 4). Compared with H20, this much improved purity likely stems from three factors: (1) a larger (by about ~60%) training sample; (2) the lenses in the training sample are all well observed in DECaLS with clearly

discernible lensing features; and (3) the non-lenses in the training sample includes a large number of images from DES that have observations with comparable depth to the lenses from DES in our training sample, which significantly reduced, if not eliminated, the ResNet's bias toward images with greater depth.

4.1.2. Candidates from Deployment on DEV and COMP in BASS/MzLS

For the northern MzLS/BASS region, the *gr*-band observations have worse seeing. Given the success of the deployment in DECaLS, however, we decide to proceed with applying our trained ResNet model, without modification or retraining, to this region.

We run inference on 5.4 million cutouts centered on DEV and COMP objects, with *z*-band magnitude <20.0. The inspection of 8761 ResNet recommendations finds 29 A's, 22 B's, and 103 C's. The locations of the candidates in the sky are shown, together with the candidates found in DECaLS, in Figure 6. In total, we have identified 154 candidates, approximately 1 in 57 ResNet recommendations. As expected, the purity of the ResNet recommendations is worse than for DECaLS, but is still competitive. Keep in mind that we used the same ResNet trained for DECaLS without any modification. Furthermore, as we mentioned in Section 2, the *gr*-band seeings are 1″61 and 1″47, respectively. To our knowledge, this is the first time a lens search has been attempted and successfully carried out, with competitive neural network recommendation purity, for a survey with seeing $\gtrsim 1''5$. This is a remarkable result.

### 4.1.3. Lens Candidates from Deployment on REX in Legacy Surveys

The REX category contains an order of magnitude more objects than the DEV and COMP types combined, since most faint, extended galaxies are modeled by the REX profile (see Section 2). This category likely includes many elliptical galaxies, though the percentage is unknown.

Given the success with DEV and COMP in both DECaLS and BASS/MzLS, without modification of the model or additional training, we deploy our trained ResNet on 6.7 million cutouts centered on REX (5 million in DECaLS and 1.7 million in BASS/MzLS), with *z*-band mag <20.0. When we performed this inference run, the source extraction and typing by *The Tractor* became available for certain patches below $\delta = -32°$. These objects are included in the deployment.

In total, we have inspected 7039 (5861 in DECaLS and 1178 in BASS/MzLS) ResNet recommendations and identified 168 candidates. Of these, 156 are in DECaLS and 12 in BASS/MzLS, resulting in recommendations with purities of 1 in 38 and 1 in 98, respectively. The average purity is ~1 in 42. We have removed candidates that have already been found in DEV and COMP (these lensing systems are "discovered" again because the cutout images containing the same systems are centered on different objects this time).

In the end, we identify 15 A's (13 in DECaLS and 2 in BASS/MzLS), 7 B's (6 in DECaLS and 1 in BASS/MzLS), and 46 C's (42 in DECaLS and 4 in BASS/MzLS), for a total of 68 new candidates. The locations of the candidates in the sky are shown in Figure 8.
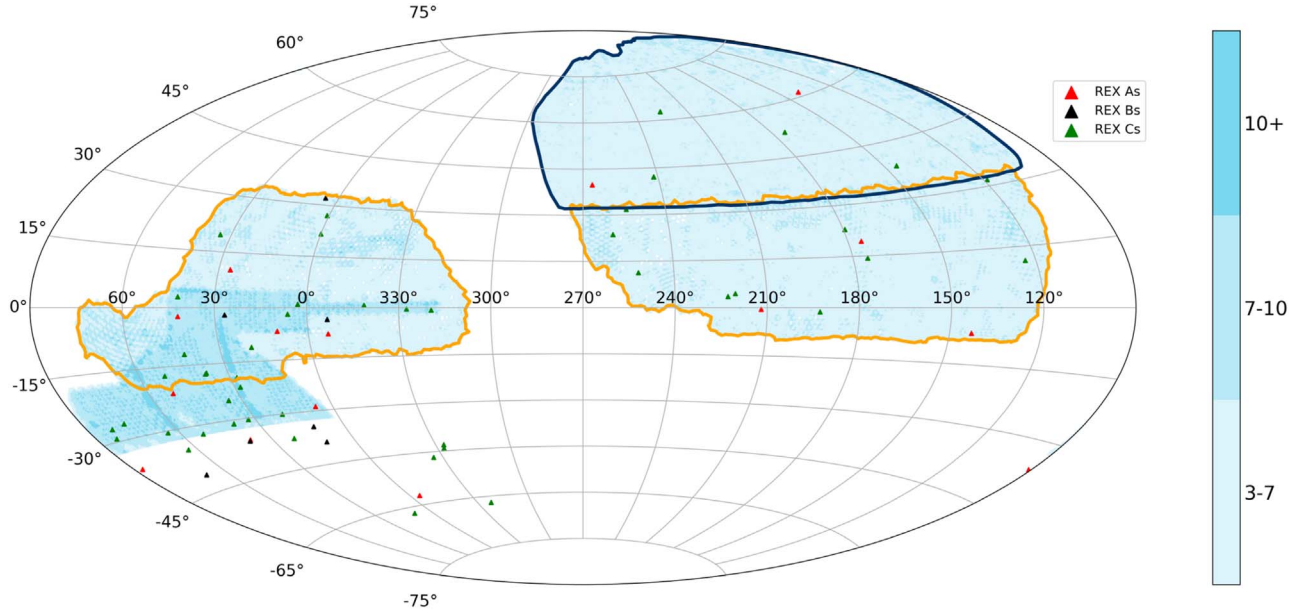
Figure 8. The new candidate lensing systems typed as REX by *The Tractor* in the Legacy Surveys are shown as red (Grade A), black (Grade B), and green (Grade C) triangles.

Table 1
L18 Model

| Grade | A | B | C | | Total by Type (DECaLS, MzLS) |
Human Score	$\geqslant 3.5$	3.0	2.5	2.0	
DC (DECaLS, MzLS)	144 (115, 29)	132 (110, 22)	280 (242, 38)	324 (259, 65)	880 (726, 154)
REX (DECaLS, MzLS)	15 (13, 2)	7 (6, 1)	22 (20, 2)	24 (22, 2)	68 (61, 7)
Total by Grade (DECaLS, MzLS)	159 (128, 31)	139 (116, 23)	302 (262, 40)	348 (281, 67)	948 (787, 161)

Table 2
Shielded Model

| Grade | A | B | C | | Total by Type (DECaLS, MzLS) |
Human Score	$\geqslant 3.5$	3.0	2.5	2.0	
DC (DECaLS, MzLS)	19 (16, 3)	20 (19, 1)	36 (34, 2)	50 (45, 5)	125 (114, 11)
REX (DECaLS, MzLS)	38 (34, 4)	40 (37, 3)	49 (47, 2)	112 (109, 3)	239 (227, 12)
Total by Grade (DECaLS, MzLS)	57 (50, 7)	60 (56, 4)	85 (81, 4)	162 (154, 8)	364 (341, 23)

All the lens candidates found by the L18 model are summarized in Table 1 (DC denotes DEV or COMP).

4.2. Candidates Found with the "Shielded" Model in Legacy Surveys

As mentioned in Section 3.3, we have experimented with modifications on the L18 ResNet model to optimize performance, but so far using the same training sample (although we will experiment with the makeup of the training sample as well). Here we present the lens candidates found by one of these attempts.

We deploy the "shielded" model on the entire Legacy Surveys footprint on objects that satisfy the same criteria as for the L18 ResNet model. We achieve a similar level of purity, and have found 364 *new* lens candidates, comprising 57 A's, 60 B's, and 247 C's. This demonstrates that a different neural network is capable of finding new lenses in the same footprint.

These lens candidates are summarized in Table 2 with their locations on the sky shown in Figure 9.

4.3. Summary of Section 4

Altogether, we have found 1312 strong lens candidates (Table 3). Of these, 102 have been found by other groups, none of which were included in our training sample. This leaves 1210 new lens candidates. Of these, there are 193 A's, 175 B's, and 842 C's. For each candidate system, we report the average numerical scores from A.D. and X.H. and the absolute difference, the region where it is found, its type from *The Tractor*, and the neural network model used. The strong lens candidates discovered in this work are summarized in Table 3. We have checked our candidate list against the spectroscopic database from the Sloan Digital Sky Survey (SDSS) DR16[19] and found that for approximately half of them the putative

---

[19] https://www.sdss.org/dr16/
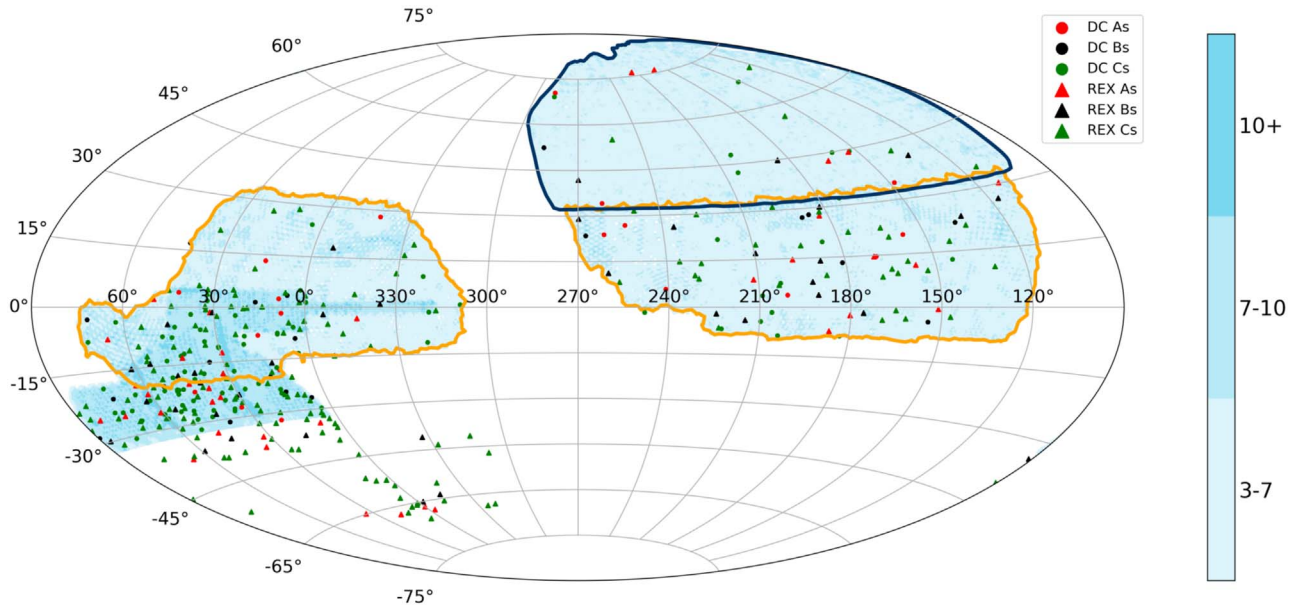
Figure 9. The new candidate lensing systems found by the "shielded" model are shown as red (Grade A), black (Grade B), and green (Grade C) circles (DEV or COMP) and triangles (REX).

Table 3
Lens Candidates

Grade	A	B	C		Total
Human Score	≥3.5	3.0	2.5	2.0	
L18 + shielded models	216	199	387	510	1312
Known lenses or candidates	23	24	27	28	102
New lens candidates in this work	193	175	360	482	1210

lensing galaxy has a spectroscopic redshift. For the rest, we have found photometric redshifts from Zhou et al. (2021).

We believe we have held a high standard in grading our candidates. Many of our Grade C systems are in fact likely lensing candidates. Among our candidates, of the 102 systems that have been identified by other groups (but were not in our training sample), 55 are in Grade C, 27 of which have a score of 2.5 (see Table 3). This speaks to the quality of our Grade C candidates. We would like to note that 42% (360) of our Grade C candidates have a human inspection score of 2.5. As shown in the examples in Figure 10 below, many of these systems are high likelihood candidates. In total, there are 728 new candidates with a score ⩾2.5.

Many of our lens candidates have spectroscopic or photometric redshifts $z \gtrsim 0.8$, greater than the typical redshifts of 0.3–0.8 for the current known lensing sample (e.g., Brownstein et al. 2012; Wong et al. 2018). In fact, the highest spectroscopic redshift from SDSS DR16 is 0.8924 (DESI-241.7346+42.1102) and the highest photometric redshift (Zhou et al. 2021) is 1.021 ± 0.061 (DESI-072.0873-19.4173). In addition, the angular scales of our systems are typically between $1\rlap{.}''5$ and $5''$ (see Figure 10), significantly larger than the typical value for previously known galaxy lensing systems ($\lesssim 1\rlap{.}''5$). This translates to longer time delays and a smaller relative uncertainty per system for quasars and supernova events in the background galaxy, and therefore higher precision in the measurement of H_0 (e.g., Suyu et al. 2020).

We end this section by highlighting in Figure 10 four examples each for four types of strong lens candidates that we have discovered. Among the 16 candidates shown, four have an average human inspection score of 2.5, and therefore are given a C grade, but they are nevertheless very likely lensing candidates.

The first 80 (sorted by R.A.) of the 216 Grade A lenses are shown in Figure 11 and Table 4. All 1312 candidates reported in this paper can be found on the website for this project[20] in this online table.[21]

5. Discussion

In our training sample there are 632 lenses. This is generally considered too small a number for training a neural network. Even our non-lens sample is much smaller than what is typically used (e.g., Jacobs et al. 2019a, 2019b). Nevertheless, we have succeeded in finding 1210 new lens candidates in the three-band Legacy Surveys with nonuniform depth (see Figure 1). The training sample was designed for searching among the DEV and COMP types in one of the two regions of the footprint, DECaLS, and our neural network model performed well for this category. The purity of our neural network recommendations is at least on par with the best in the literature. Compared with H20, using a larger training sample that includes a larger proportion of non-lenses with deep observations, we have improved the performance of our neural network model (as measured by recommendation purity) by a factor of 5 for DEV and COMP in DECaLS (from 1 in 150 to 1 in 31), where the majority of our lenses are found.

Just as significantly, our results show that our trained neural network models can be applied to data sets beyond the scope of the training sample. For DEV and COMP sources in BASS/MzLS, which has inferior gr-band seeing, and for REX sources in the entire Legacy Surveys footprint, which typically have
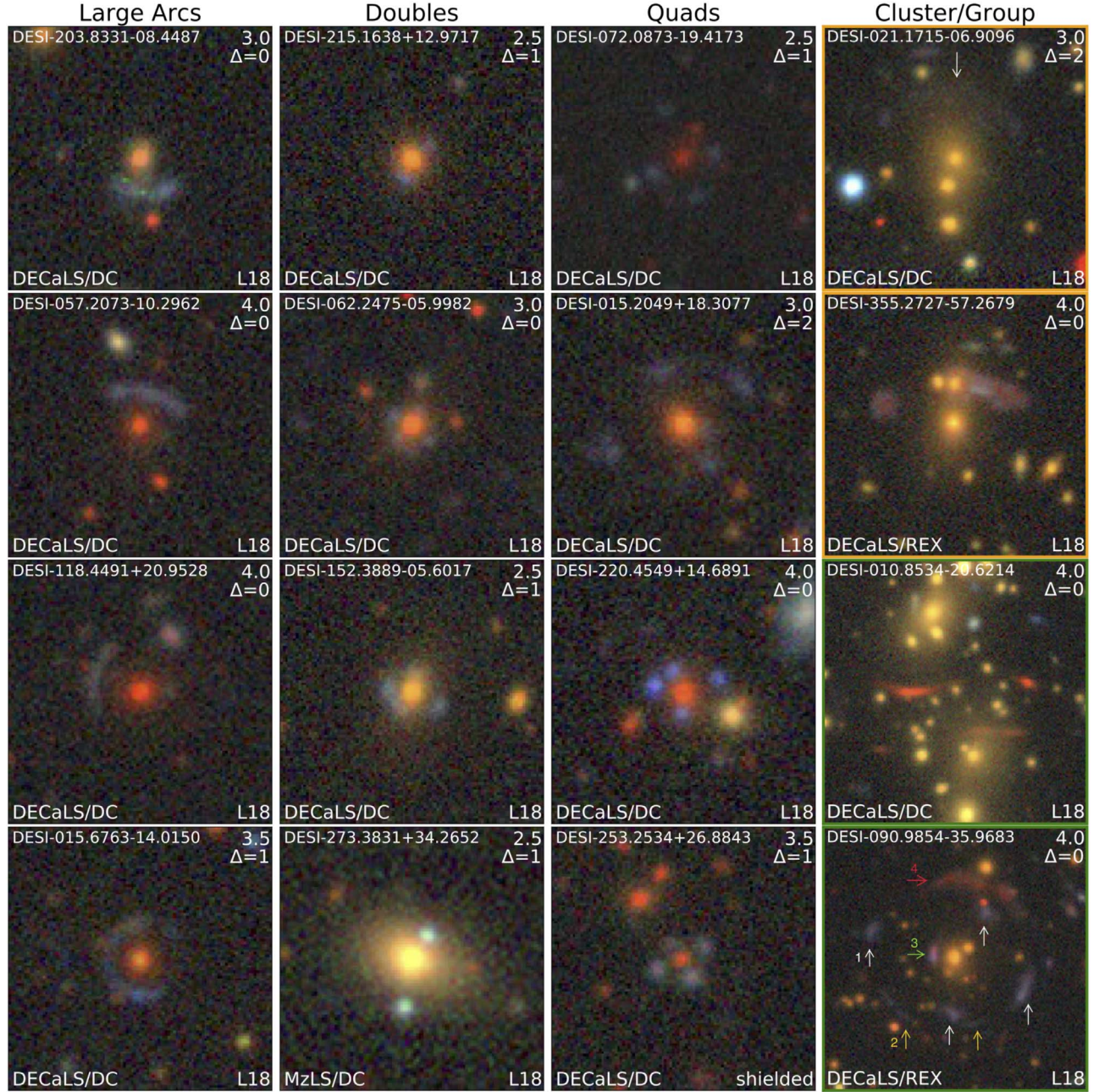
Figure 10. Sixteen of the 1210 new lensing candidates discovered in this paper. The naming convention is R.A. and decl. in decimal format. The top right corner of each image indicates the average human inspection score with Δ being the absolute difference; bottom left corner, the region and *The Tractor* type (REX or DC = DEV or COMP); and bottom right, the neural network model. North is up, and east to the left. The images without rims have a width of 101 pixels $\approx 26\rlap{.}''5$; with orange rims, 151 pixels $\approx 39\rlap{.}''6$; and green rims, 201 pixels $\approx 52\rlap{.}''7$. First column: large arcs. The third system (DESI-118.4491+20.9528) clearly has a counter-image and the fourth one (DESI-015.6763–14.0150) is a near Einstein ring. Second column: doubly lensed systems. The second (DESI-062.2475-05.9982) and third (DESI-152.3889-05.6017) systems hint at a possible Einstein cross (or a quad) and the fourth one (DESI-273.3831+34.2652) is a likely doubly lensed quasar system. Third column: quadruply lensed systems. The first system, DESI-072.0873-19.4173, has the highest photometric redshift, 1.021 ± 0.061, among our candidates. These 12 systems have a single galaxy as the main lens. Fourth column: cluster/group lensing systems. The first one (DESI-021.1715-06.9096) has a faint, giant blue arc (white arrow). The second (DESI-355.2727-57.2679) and third (DESI-010.8534–20.6214) systems show one and two sets of red arcs, respectively. The fourth one (DESI-090.9854-35.9683) is a spectacular system: at least four lensed sources at different redshifts are apparent, including a quad (1, white arrows), a "broken" long arc (2, yellow arrows), one red arc near the core of the group (3, green arrow), and a giant red arc at approximately $14''$ away from the lens center (4, red arrow). Note that the four candidates receiving a score of 2.5, and therefore a grade of C, are nevertheless very likely lensing candidates.

smaller apparent sizes, we applied the exact same trained model, and the purities are only slightly lower: 1 in 57 and 1 in 42, respectively.

To the best of our knowledge, most of the lensing systems confirmed to date are from spectroscopic searches in the SDSS (e.g., Bolton et al. 2008; Treu et al. 2011; Brownstein et al. 2012; Shu et al. 2017). These systems typically have an angular scale of $\sim 1''$. Our candidates, discovered from ground-based imaging surveys, have angular scales $\gtrsim 1''$. Thus they not only significantly expand the number of lensing systems but are complementary to most of the confirmed systems. We will leave a detailed comparison to a future publication, between the confirmed lensing systems and results from our next search.
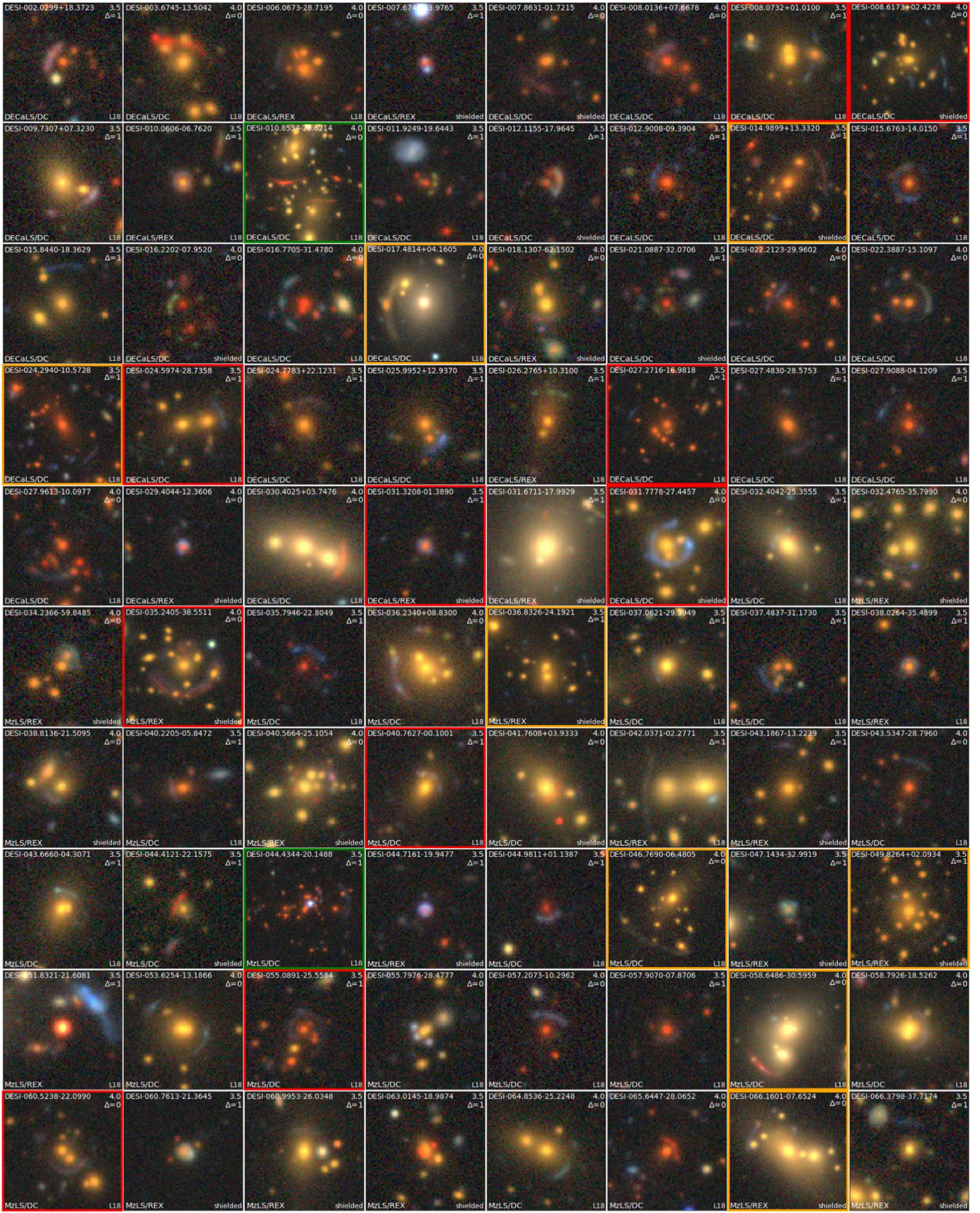
Figure 11. Eighty of the 216 Grade A candidates arranged in ascending R.A. Top right corner indicates the average human inspection score with Δ being the absolute difference; bottom left corner, the region and *The Tractor* type (REX or DC = DEV or COMP); and bottom right, the neural network model. For each image, north is up, and east to the left. Images without rims have a width of 101 pixels (26″5); with orange rims, 151 pixels (39″6); and green rims, 201 pixels (52″7). Images with red rims are known lenses or candidates but not included in our training sample, with citations given in Table 4. All 1312 candidates are shown on the project website: https://sites.google.com/usfca.edu/neuralens.

Table 4
Grade A Candidates

Name	Type	mag_g	mag_r	mag_z	Probability	z_{spec}	z_{phot}
DESI-002.0299+18.3723	DC	20.63	19.16	17.84	0.97	0.4724	
DESI-003.6745-13.5042	DC	20.00	18.17	17.18	0.34		0.406 ± 0.021
DESI-006.0673−28.7195	REX	20.80	18.86	17.55	0.97		0.533 ± 0.025
DESI-007.6741−33.9765	REX	21.64	20.85	19.61	0.20		0.858 ± 0.041
DESI-007.8631-01.7215	DC	21.61	19.79	18.60	0.53	0.5201	
DESI-008.0136+07.6678	DC	21.14	19.57	18.10	0.99	0.5566	
DESI-008.0732+01.0100 (a)	DC	20.25	18.4	17.45	0.31		0.426 ± 0.091
DESI-008.6173+02.4228 (b), (c)	DC	22.62	20.73	19.81	0.54	0.4548	
DESI-009.7307+07.3230	DC	18.65	17.21	16.36	0.22	0.2547	
DESI-010.0606−06.7620	REX	20.98	19.69	18.64	1.00	0.6152	
DESI-010.8534−20.6214	DC	18.93	17.4	16.56	0.11	0.3381	
DESI-011.9249−19.6443	DC	21.78	20.32	18.92	0.84		0.688 ± 0.056
DESI-012.1155−17.9645	DC	22.57	21.17	19.76	0.90		0.735 ± 0.069
DESI-012.9008−09.3904	DC	22.30	20.33	18.64	1.00	0.4485	
DESI-014.9899+13.3320	DC	20.07	18.09	16.86	0.91	0.5163	
DESI-015.6763−14.0150	DC	22.07	20.27	18.79	1.00		0.658 ± 0.036
DESI-015.8440−18.3629	DC	20.42	18.71	17.83	1.00		0.364 ± 0.019
DESI-016.2202−07.9520	DC	23.36	21.16	19.29	0.48		0.764 ± 0.032
DESI-016.7705−31.4780	DC	21.77	19.95	18.12	1.00		0.772 ± 0.018
DESI-017.4814+04.1605	DC	19.66	18.53	17.85	0.19		0.281 ± 0.140
DESI-018.1307−62.1502	REX	20.07	18.27	17.32	0.98		0.427 ± 0.024
DESI-021.0887−32.0706	DC	22.59	20.88	19.25	0.82		0.754 ± 0.080
DESI-022.2123−29.9602	DC	21.73	19.98	18.50	1.00		0.649 ± 0.023
DESI-022.3887−15.1097	DC	21.78	20.01	18.81	1.00		0.524 ± 0.063
DESI-024.2940−10.5728	DC	20.89	18.92	17.30	0.87	0.4135	
DESI-024.5974−28.7358 (c)	DC	20.42	18.60	17.62	1.00		0.414 ± 0.022
DESI-024.7783+22.1231	DC	21.06	19.09	17.93	0.13	0.4684	
DESI-025.9952+12.9370	DC	20.75	18.86	17.80	0.92	0.5114	
DESI-026.2765+10.3100	REX	21.56	20.16	19.51	0.28		0.462 ± 0.138
DESI-027.2716−16.9818 (c)	DC	20.91	19.07	17.65	0.92	0.6916	
DESI-027.4830−28.5753	DC	21.32	19.43	18.14	0.97		0.557 ± 0.008
DESI-027.9088−04.1209	DC	21.89	20.11	18.58	0.81	0.6412	
DESI-027.9613−10.0977	DC	22.05	20.36	18.87	1.00	0.5234	
DESI-029.4044−12.3606	REX	20.77	20.19	19.20	0.95		0.747 ± 0.169
DESI-030.4025+03.7476	DC	19.16	17.93	17.16	0.64	0.1696	
DESI-031.3208−01.3890 (c)	REX	21.01	20.18	19.05	1.00	0.6992	
DESI-031.6711−17.9929	REX	20.49	19.62	19.27	0.10	0.1548	
DESI-031.7778−27.4457 (c)	DC	18.78	17.26	16.37	1.00		0.294 ± 0.018
DESI-032.4042-25.3555	DC	18.15	16.87	16.13	0.49		0.224 ± 0.015
DESI-032.4765−35.7990	REX	21.42	19.78	18.95	0.98		0.380 ± 0.076
DESI-034.2366−59.8485	REX	21.01	19.34	18.29	0.66		0.498 ± 0.017
DESI-035.2405−38.5511 (c)	REX	18.98	17.32	16.39	0.45		0.407 ± 0.038
DESI-035.7946−22.8049	DC	22.44	21.10	19.33	1.00		0.890 ± 0.048
DESI-036.2340+08.8300	DC	18.15	16.50	15.61	1.00		0.326 ± 0.035
DESI-036.8326−24.1921	REX	19.27	17.43	16.46	0.39		0.405 ± 0.022
DESI-037.0621−29.3949	DC	19.74	18.11	17.25	0.93		0.312 ± 0.013
DESI-037.4837−31.1730	DC	22.06	20.28	19.17	0.99		0.466 ± 0.075
DESI-038.0264−35.4899	REX	21.31	20.00	18.90	1.00		0.550 ± 0.097
DESI-038.8136−21.5095	REX	22.15	20.64	19.83	0.94		0.328 ± 0.035
DESI-040.2205−05.8472	DC	21.68	19.81	18.52	0.97	0.5238	
DESI-040.5664−25.1054	REX	21.68	20.12	19.31	0.82		0.325 ± 0.045
DESI-040.7627−00.1001 (c)	DC	19.95	18.14	17.19	1.00	0.4127	
DESI-041.7608+03.9333	DC	18.53	16.99	16.14	0.30	0.2603	
DESI-042.0371−02.2771	REX	20.04	18.66	17.90	0.93		0.247 ± 0.018
DESI-043.1867−13.2239	REX	22.31	20.58	19.67	0.67		0.431 ± 0.033
DESI-043.5347−28.7960	DC	22.19	20.30	18.82	0.93		0.632 ± 0.021
DESI-043.6660−04.3071	DC	19.35	17.68	16.79	1.00	0.2880	
DESI-044.4121−22.1575	DC	21.01	19.36	18.49	0.46		0.484 ± 0.070
DESI-044.4344−20.1488	DC	23.07	21.09	19.42	0.49		0.711 ± 0.025
DESI-044.7161−19.9477	REX	20.04	19.79	19.07	0.89		0.535 ± 0.327
DESI-044.9811+01.1387	DC	21.54	20.39	18.66	1.00		0.834 ± 0.079
DESI-046.7690−06.4805	DC	22.11	20.70	19.80	0.32		0.480 ± 0.105
DESI-047.1434−32.9919	REX	20.70	19.76	19.24	0.22		0.500 ± 0.091
DESI-049.8264+02.0934	REX	22.33	20.35	19.37	0.90	0.3189	

Table 4
(Continued)

Name	Type	mag_g	mag_r	mag_z	Probability	z_{spec}	z_{phot}
DESI-051.8321–21.6081	REX	19.43	18	16.69	0.16		0.502 ± 0.201
DESI-053.6254–13.1866	DC	19.51	17.72	16.79	1.00		0.362 ± 0.010
DESI-055.0891–25.5584 (c)	DC	21.94	20.17	18.70	1.00		0.667 ± 0.047
DESI-055.7976–28.4777	DC	20.93	19.27	18.38	1.00		0.447 ± 0.049
DESI-057.2073–10.2962	DC	21.97	20.33	18.65	1.00		0.745 ± 0.075
DESI-057.9070–07.8706	DC	21.98	20.28	18.72	1.00		0.677 ± 0.051
DESI-058.6486–30.5959	DC	16.63	15.50	14.74	0.12		0.173 ± 0.021
DESI-058.7926–18.5262	DC	18.30	16.76	15.93	1.00		0.282 ± 0.011
DESI-060.5238–22.0990 (c)	DC	21.11	19.32	18.25	1.00		0.367 ± 0.051
DESI-060.7613–21.3645	REX	20.19	19.34	18.83	0.28		0.455 ± 0.100
DESI-060.9953–26.0348	REX	20.63	19.74	19.12	0.80		0.289 ± 0.083
DESI-063.0145–18.9874	REX	22.06	20.27	19.30	0.18		0.572 ± 0.073
DESI-064.8536–25.2248	DC	19.51	17.72	16.82	0.98		0.360 ± 0.012
DESI-065.6447–28.0652	DC	21.63	19.81	18.31	0.53		0.641 ± 0.036
DESI-066.1601–07.6524	REX	21.62	20.45	19.71	0.79		0.278 ± 0.049
DESI-066.3798–37.7174	REX	20.79	19.16	18.32	0.23		0.290 ± 0.041

Note Eighty of the 216 Grade A lens candidates are listed in this table. All 1312 candidates are shown on the project website: https://sites.google.com/usfca.edu/neuralens. The spectroscopic redshifts are from SDSS DR16, all with uncertainties $<3.7 \times 10^{-4}$. References for known lenses or candidates are as follows: (a) Jaelani et al. (2020), (b) Carrasco et al. (2017), (c) Jacobs et al. (2019b).
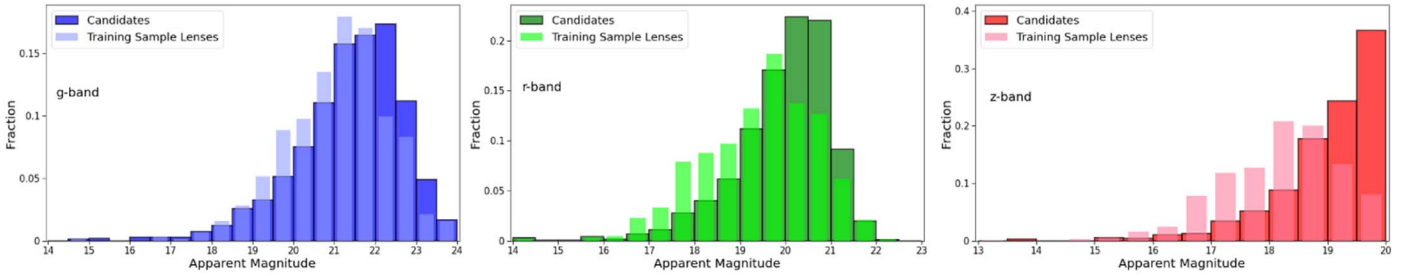
Figure 12. Magnitude distributions for *grz* bands of our candidate lensing galaxies and lenses in our training sample.
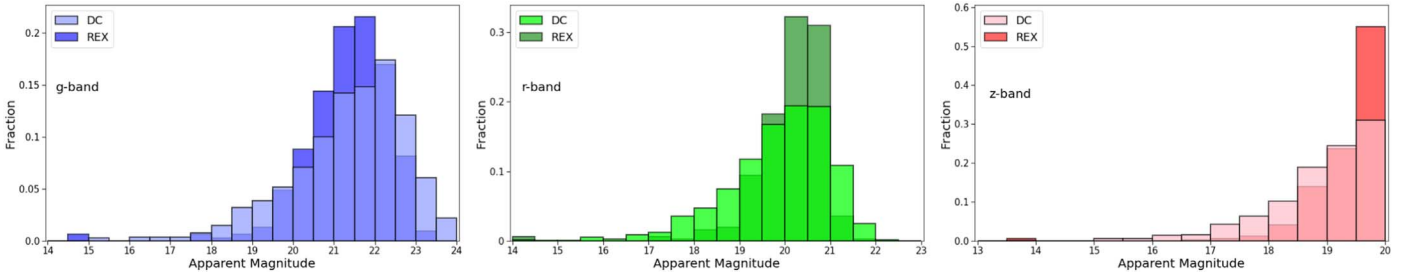
Figure 13. Apparent magnitude distributions for DEV and COMP (DC) and REX for *grz* bands.

5.1. Magnitude Distributions

For a comparison in apparent magnitudes, Figure 12 shows the distributions of the *grz* magnitudes for the candidate lensing galaxies and lenses in the training sample. Recall that for both we have imposed a cut of *z*-band mag <20.0. Fractionally, there are more brighter lenses in the training sample. This is in part because the brighter lenses are more likely to have been discovered before and are included in the training sample. Beyond this trivial difference, it is important to note the following. (1) The distributions of our candidates are fainter than the lenses in the training sample, that is, we can discover lenses that have a fainter distribution than the training lenses. (2) There are more lenses to be found fainter than *z*-band mag = 20.0 (Figure 12, third panel). In our next search we will

use a fainter magnitude limit. We will also add to our training sample lenses that are fainter than this limit, though this number is likely to be small. However, we do not expect this to be a major limitation to our ability to find fainter lenses, for two reasons: (a) we have shown that we can find lens candidates with a fainter distribution than training lenses, and (b) we will include the lens candidates found in this paper, which will greatly boost the number of lenses near the faint end of the magnitude distributions and likely will help significantly with discovering even fainter lenses. We expect to find many more lenses fainter than *z*-band mag <20.0. With the expectation that the recommendation purity of the neural net is likely to be worse for these fainter systems, we are developing new neural network models for our next search, as mentioned before. We
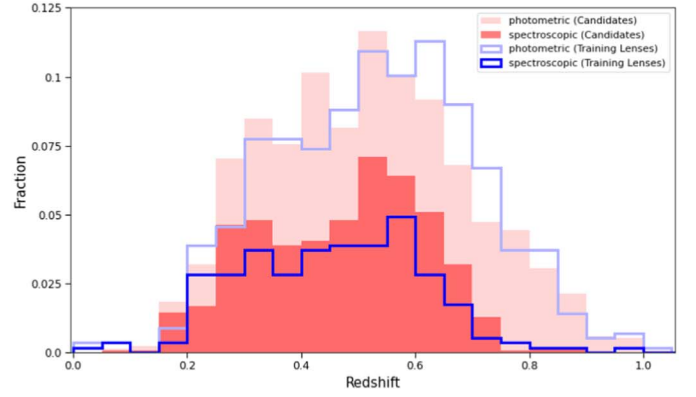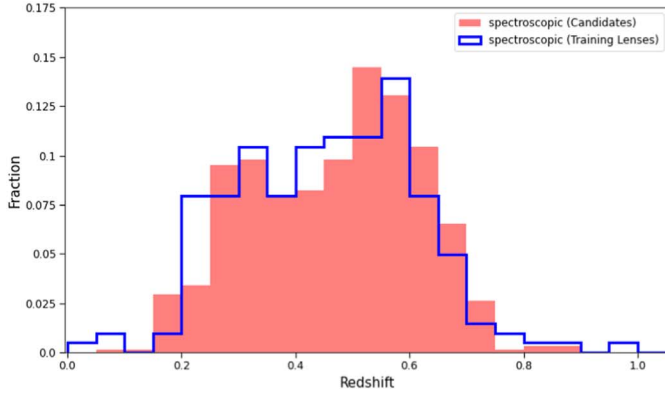
Figure 14. Distributions of the 642 of the 1312 candidates and 201 of the 632 training lenses that have spectroscopic redshifts from SDSS DR16 are shown in the left panel. The right panel adds photometric redshifts from Zhou et al. (2021) for those objects without spectroscopic redshifts.

may also carry out experiments using a fainter magnitude cut for the search than the training lenses, to see how the purity and completeness of the neural net depend on the search magnitude cut.

A more detailed comparison by galaxy type is presented in Figure 13, for DEV+COMP (DC) and REX. The REX type has a narrower distribution for all three bands. Recall that in the classification scheme of *The Tractor*, REX is the designation for those objects classified as galaxies that cannot be confidently promoted to DEV, COMP, or EXP. They are generally fainter and smaller in apparent size. This explains the smaller fractions for REX on the bright side of the distributions. Due to their smaller apparent sizes, if they are too faint, they may not be classified by *The Tractor* as spatially extended (see Section 2) at all. Hence the dearth on the faint side of the *gr* distributions. The exception is the *z* band: there is a significant contribution of REX near the magnitude limit of 20.0.

5.2. Redshift Distributions

We now compare the redshift distributions of our candidates and the lenses in the training sample (Figure 14). For spectroscopic redshifts, our candidates have approximately the same distribution as the training lenses. One notable difference is the higher fraction of training lenses for $z_{\mathrm{lens}} \lesssim 0.25$, which is expected: a large fraction of these relatively nearby lenses have been observed spectroscopically in SDSS. A more detailed discussion about the behavior of the mid-range ($0.25 \lesssim z_{\mathrm{lens}} \lesssim 0.6$) is given below in the context of comparing the distributions of DC versus REX candidates.

With photometric redshifts added for objects without spectroscopic redshifts (right panel of Figure 14), the distributions of the candidates are again similar to the lenses in the training sample. One difference is that the peak of the distribution of the training lenses is ∼0.1 higher than that of our candidates.

To understand this difference, we show the color comparison between our candidates and the training lenses (Figure 15). The training lenses are slightly redder. A strong majority of the training lenses are from the DES region, all of which are of the DC type (see Figure 2). Given that many of our candidates that are DC are from outside the DES region (DECaLS but non-DES or MzLS), and the remainder are REX from the entire Legacy Surveys footprint, it is not a surprise that there is a
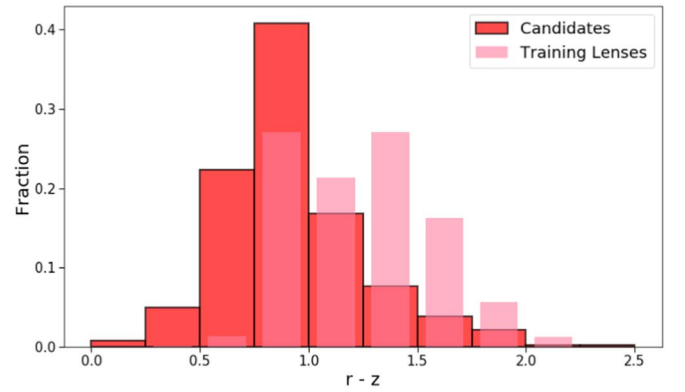
Figure 15. The $r - z$ color distributions of our candidates and the training lenses.

small color difference. The vast majority of the DES lenses in the training sample above $z = 0.5$ only have photometric redshifts. This small offset in redshift distributions is consistent with the color difference between the training lenses and our candidates.

The comparisons of redshift distributions for DC and REX are shown in Figure 16. There is clear contrast between these two types. For both spectroscopic and photometric redshift distributions, around $z_{\mathrm{lens}} \sim 0.5$, while the DC lenses peak, we find a "valley" for the REX.

For spectroscopic redshift distributions, as stated earlier, in the low redshift range ($z_{\mathrm{lens}} \lesssim 0.25$), a large percentage of DC lenses are known systems and have been included in the training sample, leaving fewer new DC lenses to be discovered. In the mid-range ($0.25 \lesssim z_{\mathrm{lens}} \lesssim 0.6$), the DC fractions are higher. For background sources at $z_s \sim 2$, the integrated optical depth for strong lensing significantly increases in this redshift range (e.g., Robertson et al. 2020). Indeed, most of the training lenses with spectroscopic redshifts are in this range (Figure 14, left panel). Therefore it is reasonable to expect that our ResNet models perform well in finding DC lenses (possibly with a high level of completeness, although, as mentioned before, it is difficult to assess completeness at this time). That we have found a large number of DC lenses is consistent with this expectation. As for the "valley" for REX, one possible explanation is that given the large number of DC training lenses in this redshift range, our neural net models are highly
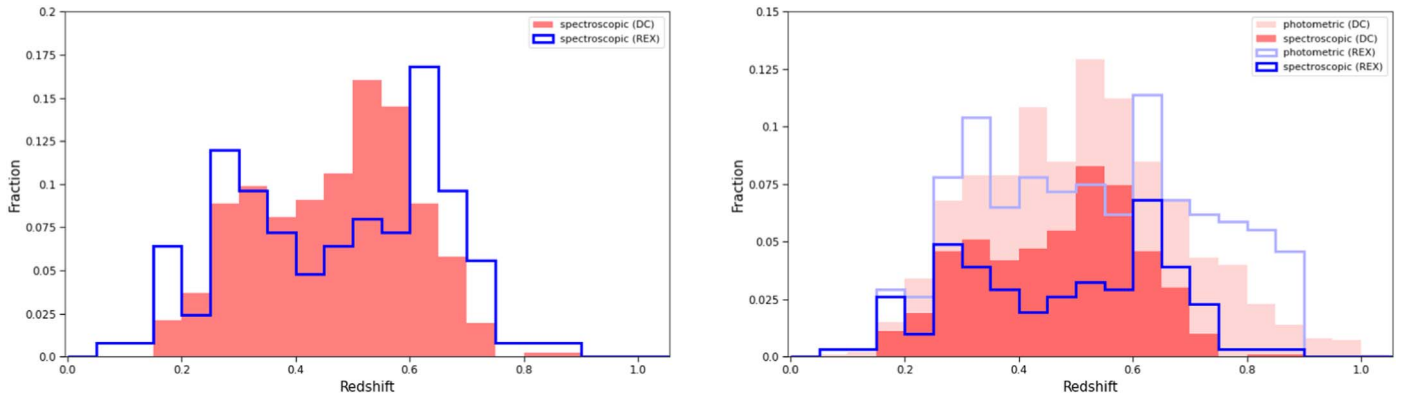
Figure 16. Distributions of the 517 of the 1005 DEV and COMP (DC) and 125 of the 307 REX candidates that have spectroscopic redshifts from SDSS DR16 are shown in the left panel. The right panel adds photometric redshifts from Zhou et al. (2021) for those objects without spectroscopic redshifts.

"tuned" to find DC lenses, and consequently somewhat biased against finding REX lens candidates. We will test this hypothesis by including REX lenses in a training sample for our next search. If true, this would suggest that there are more REX lenses to be found in this range. With photometric redshifts added for objects without spectroscopic redshifts (right panel of Figure 16), the same "valley" for the REX candidates in approximately the same redshift range can be seen.

5.3. Probability Distributions

We compare the neural network probabilities for DC and REX candidates in Figure 17. The difference for the three lowest probability bins is rather striking. It almost certainly means that at the probability threshold of 0.1, the level of completeness for REX candidates is much lower than for DC. In our discussion in Section 5.2, we suspected that this might be the case. We can improve the completeness for the REX candidates by lowering the threshold. But that likely implies a much larger set of images to inspect. As we have suggested in Section 5.2, a better remedy may be to include REX lenses in a future training sample, and this would also allow us to make a fairer comparison of the probability distributions of DC and REX candidates in our next search.

5.4. Implications for Future Searches

As we have noted above, for future searches, there is still room for improvement, both in terms of algorithm and the construction of the training sample. On the algorithm side, we have started experimenting with a variety of approaches (see Section 3.3). In this paper we have shown the results from one of them, the "shielded" model, with 364 *new* candidates found in the same footprint. This is a promising sign that further exploration is warranted. With the candidates reported in this paper and other recent discoveries (e.g., Canameras et al. 2020) we can add more lenses and lens candidates to our training sample. It is possible that further increasing the number of non-lenses in our training sample would help as well, because the current number of 21,000 is still relatively small.

We also note that we did not include any lensed QSO systems in our training sample. It is therefore not a surprise that the overwhelming majority of the candidates are galaxy–galaxy lensing systems. However, we have found a small number of lensed QSO candidate systems, e.g., DESI-273.3831+34.2652 (see Figure 10) and DESI-055.7976−28.4777 (#68 in our online
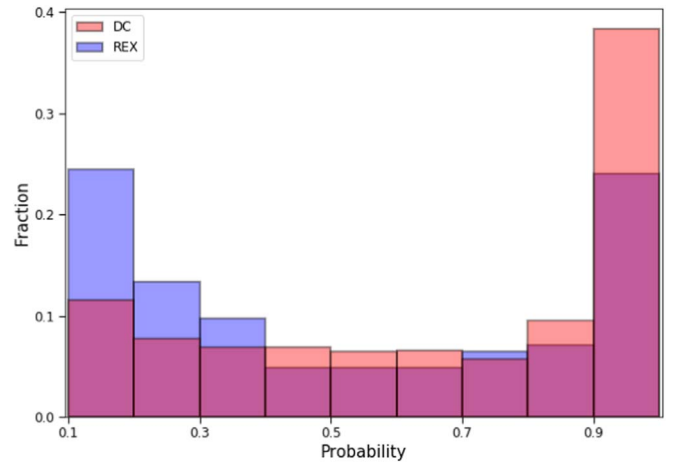
Figure 17. The probability distributions of our DEV+COMP (DC) and REX candidates.

table (see footnote 21)). In a future search, we will explore the possibility of including known lensed QSOs in the training sample, or possibly do a separate search for lensed QSOs.

The Rubin Observatory Legacy Survey of Space and Time (LSST) will commence in the near future, with $\mathcal{O}(10^5)$ strong lenses expected to be discovered (e.g., Collett 2015). We have shown that a large number of strong lenses can be discovered by using 632 observed lenses in the training sample. We have also shown that it is possible to discover lenses that are fainter than the ones in the training sample. Therefore from our experience, using observed images is a viable path for LSST. One strategy would be to find new lenses iteratively as the observation depth increases over time. As we pointed out at the end of Section 4.1.1, from the two searches we have conducted in H20 and this paper, respectively, a larger number of lenses in the training sample is likely a major factor in the improvement of the ResNet's performance. Thus this kind of "active learning" strategy will possibly result in the improvement of the neural net model from each successive training, as the observation depth increases and more lensing systems become discoverable, precisely when higher neural net purity becomes more desirable.

6. Conclusions

We have carried out a search for strong gravitational lensing systems in the DESI Legacy Surveys data by using a deep ResNet, developed by Lanusse et al. (2018), trained on

observed lenses and non-lenses. We applied our trained neural network to a total of ~20 million cutout images in DR8 with at least three passes in each of the *grz* bands and a *z*-band magnitude cut of 20.0 for the galaxy at the center of each image. We have found 193 Grade A, 175 Grade B, and 842 Grade C new candidates. These include 364 candidates found by applying a modified neural network to the same data set. We believe we have held a high standard in grading these candidate systems. 728 of our candidates have a human inspection score $\geqslant 2.5$, all of which are at least likely lensing systems.

We note that the candidates reported in this paper do not include the 335 strong lensing candidates (with 159 Grade A's and B's) we already found for Legacy Surveys DR7 (Huang et al. 2020).

Compared with efforts by other groups to search for strong lensing systems in other surveys, we use a much smaller training sample of 632 lenses and ~21,000 non-lenses from observed data, for a survey that covers one third of the sky with nonuniform depth and seeing. We nevertheless have achieved competitive neural network recommendation purity and in this paper we report the discovery of 1210 new strong lens candidates.

For our future searches, it is important to note that on one hand, our neural network models *are* capable of finding lenses outside the scope of the training sample in multiple aspects, and on the other hand, it is also vital to build a "statistically representative" training sample as much as possible to achieve a good balance between purity and completeness, so as to maximize the reward for human inspection. With the lenses we have discovered in this paper, we can now build a training sample that is more statistically representative of the Legacy Surveys for our next search.

Thus for future ground-based large surveys, such as the LSST, we suggest an iterative search strategy using observed images with an increasing number of lenses in the training sample. This "active learning" strategy will allow for discovering lenses not represented (or not well represented) in the training sample and the continual improvement of the training sample to increase the capacity and enhance the performance of the neural network model.

ORCID iDs

X. Huang ◉ https://orcid.org/0000-0001-8156-0330
C. Storfer ◉ https://orcid.org/0000-0002-0385-0014
A. Dey ◉ https://orcid.org/0000-0002-4928-4003
M. Landriau ◉ https://orcid.org/0000-0003-1838-8528
D. Lang ◉ https://orcid.org/0000-0002-1172-0754
A. Meisner ◉ https://orcid.org/0000-0002-1125-7384
J. Moustakas ◉ https://orcid.org/0000-0002-2733-4559
E. F. Schlafly ◉ https://orcid.org/0000-0002-3569-7421
D. J. Schlegel ◉ https://orcid.org/0000-0002-5042-5088

References

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017, Natur, 551, 85
Abbott, T. M. C., Abdalla, F. B., Annis, J., et al. 2018, MNRAS, 480, 3879
Auger, M. W., Treu, T., Bolton, A. S., et al. 2010, ApJ, 724, 511
Bolton, A. S., Burles, S., Koopmans, L. V. E., et al. 2008, ApJ, 682, 964
Bolton, A. S., Burles, S., Koopmans, L. V. E., Treu, T., & Moustakas, L. A. 2006, ApJ, 638, 703
Brownstein, J. R., Bolton, A. S., Schlegel, D. J., et al. 2012, ApJ, 744, 41
Canameras, R., Schuldt, S., Suyu, S. H., et al. 2020, arXiv:2004.13048
Carrasco, M., Barrientos, L. F., Anguita, T., et al. 2017, ApJ, 834, 210
Choi, S. K., Hasselfield, M., Ho, S.-P. P., et al. 2020, JCAP, 12, 045
Collett, T., Montanari, F., & Räsänen, S. 2019, PhRvL, 123, 231101
Collett, T. E. 2015, ApJ, 811, 20
Collett, T. E., Oldham, L. J., Smith, R. J., et al. 2018, Sci, 360, 1342
Cornachione, M. A., Bolton, A. S., Shu, Y., et al. 2018, ApJ, 853, 148

Dey, A., Rabinowitz, D., Karcher, A., et al. 2016, Proc. SPIE, 9908, 99082C
Dey, A., Schlegel, D. J., Lang, D., et al. 2019, AJ, 157, 168
Diaz Rivero, A., & Dvorkin, C. 2020, PhRvD, 101, 023515
Diehl, H. T., Buckley-Geer, E. J., Lindgren, K. A., et al. 2017, ApJS, 232, 15
Flaugher, B., Diehl, H. T., Honscheid, K., et al. 2015, AJ, 150, 150
Freedman, W. L., Madore, B. F., Hatt, D., et al. 2019, ApJ, 882, 34
Freedman, W. L., Madore, B. F., Hoyt, T., et al. 2020, ApJ, 891, 57
Goldstein, D. A., & Nugent, P. E. 2017, ApJL, 834, L5
Goldstein, D. A., Nugent, P. E., & Goobar, A. 2019, ApJS, 243, 6
Goldstein, D. A., Nugent, P. E., Kasen, D. N., & Collett, T. E. 2018, ApJ, 855, 22
Goobar, A., Amanullah, R., Kulkarni, S. R., et al. 2017, Sci, 356, 291
Huang, X., Storfer, C., Ravi, V., et al. 2020, ApJ, 894, 78
Jacobs, C., Collett, T., Glazebrook, K., et al. 2019a, MNRAS, 484, 5330
Jacobs, C., Collett, T., Glazebrook, K., et al. 2019b, ApJS, 243, 17
Jacobs, C., Glazebrook, K., Collett, T., More, A., & McCarthy, C. 2017, MNRAS, 471, 167
Jaelani, A. T., More, A., Oguri, M., et al. 2020, MNRAS, 495, 1291
Jauzac, M., Harvey, D., & Massey, R. 2018, MNRAS, 477, 4046
Kelly, P. L., Filippenko, A. V., Burke, D. L., et al. 2015, Sci, 347, 1459
Khetan, N., Izzo, L., Branchesi, M., et al. 2020, arXiv:2008.07754
Kneib, J.-P., & Natarajan, P. 2011, A&ARv, 19, 47
Kochanek, C. S. 1991, ApJ, 373, 354
Koopmans, L. V. E., & Treu, T. 2002, ApJL, 568, L5
Koopmans, L. V. E., Treu, T., Bolton, A. S., Burles, S., & Moustakas, L. A. 2006, ApJ, 649, 599
Lang, D., Hogg, D. W., & Mykytyn, D. 2016, The Tractor: Probabilistic Astronomical Source Detection and Measurement v. dr7.0, Astrophysics Source Code Library, ascl:1604.008
Lanusse, F., Ma, Q., Li, N., et al. 2018, MNRAS, 473, 3895
Li, R., Napolitano, N. R., Tortora, C., et al. 2020, ApJ, 899, 30
Li, Z.-X., Gao, H., Ding, X.-H., Wang, G.-J., & Zhang, B. 2018, NatCo, 9, 3833
Linder, E. V. 2011, PhRvD, 84, 123529
Marshall, P. J., Treu, T., Melbourne, J., et al. 2007, ApJ, 671, 1196
Meneghetti, M., Davoli, G., Bergamini, P., et al. 2020, Sci, 369, 1347
Metcalf, R. B., Meneghetti, M., Avestruz, C., et al. 2019, A&A, 625, 119
Monna, A., Seitz, S., Balestra, I., et al. 2017, MNRAS, 466, 4094
Moustakas, L. A., Brownstein, J., Fadely, R., et al. 2012, AAS Meeting, 219, 146.01
Narayan, R., & Bartelmann, M. 1996, arXiv:astro-ph/9606001
Nightingale, D. R., Massey, J. W., Harvey, R. J., et al. 2019, MNRAS, 489, 2049

Oguri, M., & Marshall, P. J. 2010, MNRAS, 405, 2579
Patrício, V., Richard, J., Carton, D., et al. 2019, MNRAS, 489, 224
Petrillo, C. E., Tortora, C., Chatterjee, S., et al. 2017, MNRAS, 472, 1129
Petrillo, C. E., Tortora, C., Vernardos, G., et al. 2019, MNRAS, 484, 3879
Philcox, O. H. E., Sherwin, B. D., Farren, G. S., & Baxter, E. J. 2020, arXiv:2008.08084
Pierel, J. D. R., & Rodney, S. 2019, ApJ, 876, 107
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, A&A, 641, A6
Pourrahmani, M., Nayyeri, H., & Cooray, A. 2018, ApJ, 856, 68
Quimby, R. M., Oguri, M., More, A., et al. 2014, Sci, 344, 396
Räsänen, S., Bolejko, K., & Finoguenov, A. 2015, PhRvL, 115, 101301
Refsdal, S. 1964, MNRAS, 128, 307
Riess, A. G., Casertano, S., Yuan, W., Macri, L. M., & Scolnic, D. 2019, ApJ, 876, 85
Ritondale, E., Vegetti, S., Despali, G., et al. 2019, MNRAS, 485, 2179
Robertson, A., Smith, G. P., Massey, R., et al. 2020, MNRAS, 495, 3727
Rodney, S. A., Strolger, L. G., Kelly, P. L., et al. 2016, ApJ, 820, 50
Shajib, A. J., Birrer, S., Treu, T., et al. 2019, MNRAS, 483, 5649
Shajib, A. J., Treu, T., Birrer, S., & Sonnenfeld, A. 2020, arXiv:2008.11724
Shu, Y., Brownstein, J. R., Bolton, A. S., et al. 2017, ApJ, 851, 48
Sonnenfeld, A., Chan, J. H. H., Shu, Y., et al. 2018, PASJ, 70, S29
Sonnenfeld, A., Jaelani, A. T., Chan, J., et al. 2019, A&A, 630, A71
Sonnenfeld, A., Treu, T., Marshall, P. J., et al. 2015, ApJ, 800, 94
Suyu, S. H., Huber, S., Cañameras, R., et al. 2020, A&A, 644, 162
Szegedy, C., Liu, W., Jia, Y., et al. 2014, arXiv:1409.4842
Taubenberger, S., Suyu, S. H., Komatsu, E., et al. 2019, A&A, 628, L7
Tessore, N., Bellagamba, F., & Metcalf, R. B. 2016, MNRAS, 463, 3115
Treu, T. 2010, ARA&A, 48, 87
Treu, T., Dutton, A. A., Auger, M. W., et al. 2011, MNRAS, 417, 1601
Treu, T., & Marshall, P. J. 2016, A&ARv, 24, 11
Tzeng, E., Hoffman, J., Saenko, K., & Darrell, T. 2017, arXiv:1702.05464
Vanzella, E., Meneghetti, M., Caminha, G. B., et al. 2020, MNRAS, 494, L81
Vegetti, S., Despali, G., Lovell, M. R., & Enzi, W. 2018, MNRAS, 481, 3661
Vegetti, S., & Koopmans, L. V. E. 2009, MNRAS, 400, 1583
Vegetti, S., Koopmans, L. V. E., Auger, M. W., Treu, T., & Bolton, A. S. 2014, MNRAS, 442, 2017
Verde, L., Treu, T., & Riess, A. G. 2019, NatAs, 3, 891
Williams, G. G., Olszewski, E., Lesser, M. P., & Burge, J. H. 2004, Proc. SPIE, 5492, 787
Wojtak, R., Hjorth, J., & Gall, C. 2019, MNRAS, 487, 3342
Wong, K. C., Sonnenfeld, A., Chan, J. H. H., et al. 2018, ApJ, 867, 107
Wong, K. C., Suyu, S. H., Chen, G. C. F., et al. 2020, MNRAS, 498, 1420
Zhou, R., Newman, J. A., Mao, Y.-Y., et al. 2021, MNRAS, 501, 3309